

Komló Csaba

Eszterházy Károly Főiskola

csabakom@ektf.hu

DIGITÁLIS MOZGÓKÉPEK ELEMZÉSÉNEK ALGORITMIKUS MEGKÖZELÍTÉSE

Bevezetés

Az információs társadalomban az információ- és kommunikációtechnológia fejlődésével egyre több információ vesz körül bennünket, egyes vizsgálatok szerint évenként kb. 50–80%-kal nő az emberiség által előállított elektronikus adatmennyiség, de vannak olyan kutatók, akik szerint a 100% sem túlzó becslés. Nem meglepő ezért, hogy mindennapjainkban –legyen szó akár munkáról, akár szórakozásról– információáradat vesz minket körül és ebben az információs tengerben kell megtalálnunk a számunkra fontos elemeket.

Az információs társadalom korában szinte mindenkinek szembe kell néznie azzal, hogy adott feladat elvégzéséhez online adatbázisok rekordjaiból, offline elérhető dokumentumokból, valamint internetes forrásokból származó adatokat is fel kell használnia.

Az említett adatok már jó ideje elektronikusan keletkeznek, és informatikai eszközökkel rendezzük, tároljuk, továbbítjuk, keressük és használjuk fel őket. A keresés hatékonyságát több tényező is befolyásolja, de kétségtelenül az egyik legfontosabb ezek közül az információ hordozójának médiuma: jó esélyünk van a találatra abban az esetben, ha az információ szöveges formában van, de mi a helyzet akkor, ha más médiumról, pl. mozgóképről van szó?

Mozgóképek számítógépes feldolgozása

A mozgóképek állóképek sorozatából épülnek fel, akár a hagyományos celluloid filmekre, akár a 20. század második felében elterjedt videóra gondolunk. A számítógépek számítási teljesítményének fejlődésével és a tárhely olcsóbbá válásával ma már a megszülető mozgóképek túlnyomó többsége digitális formában kerül tárolásra és a korábban használt formátumok archiválása is ebben a formában történik.

A digitális formátumban tárolt mozgóképek vagy röviden: digitális mozgóképek mérete rendkívül széles skálán mozog. A legkevesebb tárolókapacításra a VHS videokazetták archivált változatainak van szükségük. A legtöbb ilyen filmet 352×288-pixeles felbontással digitalizálják, és az 50 félkép/másodperces képrátát 25 egész képpé alakítják át. A felbontásra vonatkoztatva tömörítetlen tárolási eljárást alkalmazva ez a felbontás és a 25 kép/másodperces képráta 3,8 megabájtnyi adatot eredményez másodpercenként. Anélkül, hogy belemennénk a videotömörítési eljárások részleteibe, fontos kiemelni, hogy ennél a számításnál is alkalmaztunk tömörítést, de a színekre és nem a felbontásra vonatkozóan.

Hasonló számítási eljárások alapján a ma általánosan használt HD felvételek (1920×1080 pixel, 59.94 kép/másodperc) 187 megabájtot, míg a napjainkban az otthoni mozizásban újdonságnak számító Ultra HD (4096×2160) mozgóképek több mint 1 gigabájtot foglalnának el a tárolómédiumon tömörítés nélkül. Szerencsére a tömörítési eljárások segítségével a tárolókapacitási igények jelentősen csökkenthetők, de így is látható, hogy a mozgóképek algoritmikus elemzésénél jelenetös problémát okoz a mozgóképek tárolása és a hatalmas mennyiségű adat feldolgozásához szükséges számítási kapacitás. Fontos megjegyezni, hogy az előbbi számításoknál csak a képi információ tárolásához szükséges tárhellyel foglalkoztunk és figyelmen kívül hagytuk az auditív elemeket.

A mozgóképek algoritmikus feldolgozásának másik sarkalatos pontja, hogy míg a digitális formában tárolt szöveges dokumentumok átalakítás nélkül, természetükből adódóan hatékonyan kereshetők, addig a mozgóképek bithalmazai semmilyen információval nem szolgálnak a tartalomra nézve, éppen ezért valamilyen járulékos információval kell ellátnunk a vizsgálat szempontjából állóképek sorozataként tekintett médiumokat. Az imént említett eljárásnak az egyik módja a metaadatokkal való ellátás.

A metaadatok

Ha nagyon röviden szeretnénk megfogalmazni a metaadatok lényegét, azt mondhatnánk: a metaadatok adatok az adatokról, pontosabban információk az adatokról. Ez azt jelenti, hogy meghatározott entitásokról (estünkben képi információhordozókról) tartalmaznak különféle információkat, pl. kulcsszavakat az adott kép tartalmára vonatkozóan stb. A metaadatok segítségével a médiaelemek hatékonyan kereshetővé válnak. Többféle metaadatozási rendszer közül választhatunk, a továbbiakban a két talán legelterjedtebb metaadatozási rendszert, a Dublin Core-t és a Learning Object Metadata rendszert ismertetjük.

Dublin-Core metaadatok

A Dublin Core Metadata Initiative (DCMI) gyakran használt metaadatozási elv. 1995-ben Dublinban (Ohio állam, USA) hozták létre a webes erőforrások leírására és kategorizálására, röviden DC-vel szokták jelölni.

A meta-adatok kialakításánál négy fő szempontot vesznek figyelembe:

1. Egyszerűség: a DCMI meta-adatait bárki számára azonnal érthetővé, elsajátíthatóvá igyekeztek tervezni, ezzel elősegítve az interneten való keresést.
2. Szemantikus interoperabilitás: a DCMI olyan struktúrát dolgozott ki a meta-adatok szerkezetére, mely minden tudományág számára lehetővé teszi alkalmazhatóságát.
3. Többnyelvűség: igyekeznek minél több nyelv sajátosságait figyelembe venni, valamint a kereshetőséget, a DC-k alkalmazhatóságát lehetővé tenni.
4. Bővíthetőség: az elterjedés és a minél szélesebb körű alkalmazhatóság érdekében lehetőség van a DC elemek folyamatos bővítésére, finomítására.

A bővíthetőséget szolgálja az is, hogy a DC tulajdonképpen két rétegből áll: az ún. egyszerű DC (simple DC), valamint minősített DC (qualified DC). Az egyszerű DC 15 különböző elemből áll:

1. Cím / title
2. Létrehozó / creator
3. Közreműködő / contributor
4. Kiadó / publisher
5. Azonosító / identifier
6. Forrás / source
7. Kapcsolat / relation
8. Dátum / date
9. Formátum / format
10. Típus / type
11. Téma / subject
12. Tartalmi leírás / description
13. Tér-idő vonatkozás / coverage
14. Nyelv / language
15. Jogok / rights

LOM metaadatok

Az elektronikus médiumok körében az egyik leggyakrabban alkalmazott metaadat szabványt az IEEE (Institute of Electrical and Electronics Engineers) szervezet jegyzi, és IEEE LOM (Learning Object Metadata) néven vált ismertté. E szabvány hierarchikus struktúrája 9 kategóriát definiál. A kategóriák mindegyike sajátos szempontokat tartalmaz (pl.: általános, technikai, oktatási információk stb.). A kategóriák további elemekre bonthatóak, pl. az általános kategórián belül találjuk a kulcsszavakat.

- Általános: az entitás általános leírására szolgál;
- Életciklus: az erőforrások életciklusával kapcsolatos tulajdonságok;
- Meta-metaadat: magáról a metaadatról ad információt;
- Technikai: az erőforrások technikai jellemzi;
- Oktatási: oktatási és pedagógiai tulajdonságok;
- Tulajdonjogok: szellemi tulajdonjogok és felhasználói jogok feltételei;
- Kapcsolat: más tananyagelemekhez való kapcsolódás jellemzői;
- Megjegyzés: megjegyzések a szolgáltatások oktatási használatával kapcsolatban;
- Besorolás: a tananyagelemek kapcsolata egy adott másik besorolási rendszerhez;

Mivel a LOM metaadatozási rendszert eredetileg digitális oktatási erőforrások leírására találták ki, ezért az oktatási területen kívül csak egyes elemeit használják, így a LOM elnevezés időnként nagyon különböző, a szabványtól jelentősen eltérő metaadatozási rendszert takar.

A metaadatrendszerek rendszerint nem alkalmasak a mozgóképek teljes körű automatikus leírására. Már csak azért sem, mert a digitalizált és a digitális mozgóképek jelentős része nem tartalmaz semmilyen járulékos információt, így metaadatokat sem. Csak az utóbbi néhány évben vált lehetővé, hogy a mozgókép készítésére alkalmas

eszközök automatikusan hasznos járulékos információval lássák el a fájlokat a keletkezésükkor, mint pl.: a létrehozás ideje, a beépített GPS vevőknek köszönhetően a létrehozás helye stb. Sajnos a mozgóképek tartalmára ezek az adatok sem utalnak közvetlenül, ezért a metaadatrendszerek csak emberi közreműködéssel alkalmasak a mozgóképek tartalmának leírására, azaz valakinek meg kell néznie az adott mozgóképet és a tartalmára vonatkozó metaadatokat megadni. A metaadatokkal ellátott mozgókép egy adatbázisba bekerülve már bármikor visszakereshető a metaadatok alapján, ekkor úgy tűnhet, hogy a gépi intelligencia elemzi a mozgókép tartalmát és ennek eredménye a kívánalmaknak megfelelő találat.

Mozgóképek automatikus feldolgozása

Ahogy a bevezetőben már említettük, a hordozó médiumtól függetlenül a mozgóképek állóképek sorozatából állnak, amit az emberi beavatkozás nélküli feldolgozásnál ki is használunk: a feldolgozás során a mozgóképek helyett az állóképeket elemezzük.

Az állóképek elemzésénél az első lépcső a színek vizsgálata. Az elektronikusan megjelenített képek három színösszetevőből állnak, ezek a vörös, zöld és kék (red, green, blue vagy röviden: RGB). A színhisztogram vizsgálat során kiszámolják az egyes színcsatornák intenzitását, majd a csatornánkénti intenzitásokra alapozva egy valószínűségi értékkel látják el a három színcsatorna együttesét. A színhisztogram vizsgálat viszonylag kevés információt nyújt a képek tartalmáról, de az előnyei között meg kell említeni, hogy a csatornák intenzitásértékei függetlenek a kép orientációjától. Egyes szakirodalmak megemlítik, hogy a hatékonyság növelése érdekében a vizsgálat előtt az RGB csatornákat átkonvertálják színezet, telítettség, fényerő (HSB, Hue, Saturation, Brightness) csatornákká.

A színhisztogram vizsgálatot a hatékonyság növelésének érdekében ki szokták egészíteni egyéb vizsgálatokkal, ilyen pl. a színkorrelogram elemzése. A színkorrelogram vizsgálat során a vizsgált képen kijelölünk egy pixelt, majd megvizsgáljuk, hogy mennyi a valószínűsége annak, hogy egy adott távolságra található pixel is ugyanolyan színű, mint az eredetileg kiválasztott pixel.

Textúra

A képek automatikus elemzésének másik, gyakran alkalmazott eleme a textúra. A textúra egyszerre jelenti az anyagszerűséget és ennek a vizuális reprezentációját. Az emberi értelem számára számos textúra könnyen felismerhető és fontos vizuális információt hordoz a vizsgált objektumról (pl. kavicsok, szövet, csiszolópapír, fakéreg stb.), éppen ezért a képek automatikus elemzésénél is fontos szerepet játszik. Ahhoz, hogy a képek textúráját hatékonyan tudjuk vizsgálni, számos képmódosító eljárást alkalmazunk: elhagyjuk a színinformációt és különféle szűrőket alkalmazunk annak érdekében, hogy a texturális jellemzők minél jobban érvényesüljenek.

A textúrák vizsgálatánál számos jellemzőt figyelembe lehet venni, a leggyakrabban alkalmazott 6 jellemző az alábbi¹: szemcseméret², finomság (coarseness), kontraszt (contrast), irányultság (directionality), vonalszerűség (linelikeness), szabályosság (regularity), durvaság (roughness).

Forma (Shape)

A forma az állóképen található objektumok kiterjedésének a reprezentációja. Az emberi tudat számára a forma az egyik legfontosabb jellemző a vizuális objektumok azonosításában, éppen ezért a képek automatikus elemzésénél is nagy jelentőséggel bír. A formák számítógépes azonosítása már csak azért sem egyszerű feladat, mert egyrészt a valóságban háromdimenziós objektumok kétdimenzióssá redukálva jelennek meg az állóképeken, másrészt a kép tartalma szempontjából fontos objektumok rendszerint nem elkülönülve jelennek meg, hanem egyéb objektumok mellett, esetleg azok részleges takarásában vagy vizuális zajjal terhelt, torzított képen stb.. Ráadásul annak a meghatározása sem könnyű feladat, hogy a képen található objektumok közül melyik fontos és melyik nem az.

Ahogy a legtöbb képfeldolgozó eljárásnál, itt is beszélhetünk globális és lokális alkalmazásáról a képelemzési eljárásnak. Ebben az esetben ez azt jelenti, hogy a vizsgált kép egészén (globális) vagy csak bizonyos részein (lokális) futassuk le a formák meghatározására alkalmas algoritmusokat. Számos matematikai algoritmus létezik, amelyek megpróbálják elkülöníteni a háttértől az objektumokat, illetve az élek megkeresésével vektorok segítségével leírni az alakzatokat. Az igazsághoz hozzátartozik, hogy ezek az algoritmusok nem minden esetben működnek tökéletesen (jól elkülöníthető a háttértől a vizsgált objektum, az objektum formája jellegzetes stb.). A következő lépés, hogy az alakzatokat „felismerje” a számítógép. Természetesen nem valódi felismerésről van szó, hanem a számítógép összehasonlítja a vizsgált objektumot (pontosabban az objektumból kinyert jellemzőket) azokkal a referenciaobjektumokkal (illetve azok jellemzőivel), amelyek egyfajta vizuális szótár elemeiként kerültek eltárolásra. Nem csak azok az algoritmusok és eljárások fejlődnek, amelyek a képek tulajdonságait nyerik ki, hanem azok is, amelyek összehasonlítják a vizsgált és a mintaobjektumokat.

Li és Wang tanulmányában³, amely a képelemzési és visszakeresési algoritmusok hatékonyságát vizsgálja, és többek között bemutat néhány sikeres és kevésbé sikeres példát. A képhez automatikusan hozzárendelt fogalmak fontossági sorrendben jelennek meg.

¹ H. Tamura, S. Mori, and T. Yamawaki. Texture features corresponding to visual perception. IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-8, no. 6, 1978, 460 – 473.

² A coarseness eredeti jelentése: durvaság. Azonban itt a kifejezés arra vonatkozik, hogy mekkorák a textúrát kialakító elemi részek. A fordítás éppen ezért koránt sem tökéletes: a szemcseméret jól illusztrálja a fogalom jelentését pl. közetszerű anyagoknál, de pl. a textil- és fakéregszerű anyagokra nem vonatkoztatható. Ennek ellenére sem használnánk az eredeti elnevezést, mert az így könnyen összekeverhető lenne a hatodik jellemzővel, a durvasággal, ami az első két jellemző (coarseness és a contrast összege).

³ Jia Li, Wang, J.Z.: Real-Time Computerized Annotation of Pictures, Pattern Analysis and Machine Intelligence, IEEE Transactions on (Volume:30 , Issue: 6) 2008, február 8.

A sikeres példák:

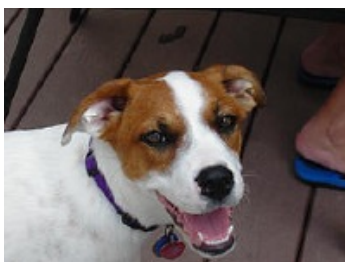


Virág, növény, rózsza, kaktusz, növényzet, fű, tájkép, élő növény

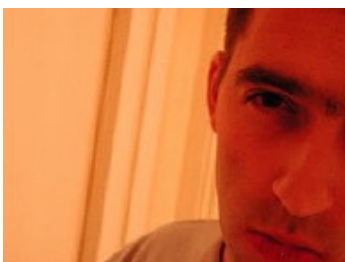


Tájkép, épület, történelmi, hegy, ember-alkotta, szobabelső, emberek, tó, állat

A sikertelen példák:



Épület, ember, víz, modern, város, munka, történelmi, ruha, ló.



Textúra, természet, virág, tenger, mikroszkópikus kép, gyümölcs, étel, zöldség, szobabelső

A tanulmány szerint a sikertelen példák okai a következők voltak: szokatlan háttér, homályos kép, nem teljesen látszó forma, hibás fényviszony.

A szemantikai szakadék

A szemantikai szakadék az emberi intelligencia és a gépi értelmezés közötti különbségből ered. Az állóképek automatikus feldolgozásánál megállapítottuk, hogy a pl. képek színösszetevőinek elemzésével a számítógép nagy valószínűséggel meg tudja „mondani”, ha egy képen a zöld szín a domináns (pl. azért, mert a képen háttérében egy zöld rét látható), illetve a formák vizsgálatával, hogy egy kör alakú objektum is van az adott képen (pl. egy labda).

Ugyanez a kép az emberi intelligencia felől megközelítve egy labda a réten, vagy sokkal inkább lehet egy futballmérkőzés utáni csendélet. Tehát a gépi intelligencia csak alapvető képelemek (színek, textúrák, egyszerűbb formák) felismerésére képes, míg az emberi tudat sokkal elvontabb módon viszonyul a vizuális információhoz. A két megközelítés közötti különbséget nevezzük szemantikai szakadéknak, amelynek csökkentése a témakörben zajló vizsgálatok elsődleges kutatási területei közé tartozik, melynek eredményeképpen a szemantikai szakadék folyamatosan csökken, de nagyon valószínű, hogy soha nem fog megszűnni.

Az MPEG 7

A mozgóképi tartalmak emberi beavatkozás nélküli elemzésén kívül szükség van egy olyan eszközre, amelyik alkalmas a mozgóképi tartalmak alacsony szintű (gépileg értelmezhető) és magasszintű (az emberi intelligenciára jellemző) leírásához elengedhetetlen specifikus jellemzők szabványszintű megadásához. Ez a szabvány 2002-ben, MPEG 7 néven született meg. Az MPEG a Motion Picture Expert Group (mozgókép szakértői csoport) az audiovizuális médiumok tömörítési eljárásával foglalkozó szakértői szervezet. Szabványait számokkal jelölik, széles körben ismert pl. az MPEG 1 tömörítési szabvány, amely elsőként definiálta a digitális mozgóképek tömörítési eljárását, ezt a szabványt használták pl. a Video CD-k előállításánál, de említhetnénk az MPEG-2 szabványt, amit többek között a DVD-nél alkalmaztak.

Az MPEG 7 szabvány azonban jelentősen különbözik a korábbi MPEG szabványoktól, mert a célja nem tömörítési eljárások leírása, hanem interfész biztosítása a multimédia tartalmak leírásához (Multimedia Content Description Interface). A szabvány részei a következők⁴:

- System: az MPEG 7 architektúrális alapjait adja meg illetve deszkriptorainak bináris formátumát. A deszkriptor egy adott tulajdonság szintaktikai és szemantikai reprezentációja.
- Description Definition Language: az MPEG 7-ben használható deszkriptorok szerkezeti kapcsolatait írja le XML nyelven. Lehetővé teszi a deszkriptorok és deszkriptor sémák létrehozását és módosítását;
- Visual: vizuális elemek leírása;
- Audio: auditív elemek leírása;
- Multimedia Descriptor Scheme: multimédia elemek leírása;

⁴ Az MPEG 7 bemutatásánál szándékosan nem kerültek lefordításra a szabvány alkotóelemeinek, hiszen az elméleti tanulmányokban és a gyakorlati alkalmazásnál is szinte minden esetben az eredeti, angol elnevezéseket használják.

- Reference software: a szabvány által leírt bináris jelsorozat előállítás
- Conformance: az alkalmazott metódusok és bitsorozatok tesztelésének specifikációja;
- Extraction and use of MPEG-7 descriptions: MPEG-7 deszkriptorok használatára vonatkozó előírások;
- Profiles: a teljes elnevezés Profiles and Levels, illetve Audiovisual Description Profile (AVDP), bevezetésének célja, hogy megkönnyítsék az automatikus információkinyerést a multimédia elemekből illetve ezek megosztását más rendszerekkel;
- Schema definition: a multimédia-tartalom leírására szolgáló metaadatrendszer definícióját tartalmazza;
- MPEG-7 profile schemas: az MPEG-7-nél alkalmazott profilok sémájának leírása;
- Query format: az audiovizuális elemek keresési mechanizmusát és szintaktikáját leíró elem
- Compact Descriptors and Visual Search: a platformfüggetlen vizuális deszkriptorok keresési hatékonyságának növelésére irányuló technológiák leírása

A deszkriptorok definiálása XML nyelven történik, Az XML (Extensible Markup Language) a W3C által ajánlott általános célú leírónyelv. Az SGML (Standard Generalized Markup Language) egyszerűsített változata, mely különböző adattípusok leírására képes. Az XML-en alapuló nyelvek (pl. MathML) leírása formális, ami lehetővé teszi a programok számára a dokumentumok módosítását és validitásának ellenőrzését a formátum előzetes ismerete nélkül (DTD, Document Type Definition). Az MPEG-7 az általános XML sémákon kívül használ néhány speciális elemet, mint pl. a mátrix adattípus.

A vizuális elemek leírása

Nem célunk az MPEG-7 nagyon részletes bemutatása, de a mozgóképek leírása szempontjából egyik legfontosabb elemről, a vizuális deszkriptorokról leírunk néhány gondolatot. Az MPEG-7 vizuális deszkriptorai az alábbi kategóriák leírására alkalmasak: szín (color), textúra (texture), alak (shape), mozgás (motion), helymeghatározás (localization), és arcfelismerés (face recognition). A kategóriák egyaránt tartalmaznak egyszerű és összetett deszkriptorokat.

Az alapdeszkriptorok a következők: grid layout, time series, multiple view, spatial 2D coordinates, temporal interpolation. Az alap deszkriptorok közül pl. a grid layout (szó szerint rácsos megjelenítésre lehetne fordítani, de a rácsra illesztés közelebb áll a fogalom tartalmához) feladata, hogy a képet egy négyzetrácsos hálóra illessze, egyértelműen meghatározhatóvá téve annak régióit.

Összefoglalás

Az információs társadalomban az exponenciálisan növekvő vizuális adatmennyiség, jelentős részének tartalmáról szinte semmilyen információnk nincs. Ezeket a fájlokat metaadatok segítségével írhatjuk le, azonban emberi erőforrással ez lehetetlen feladat.

Egyedül a számítógépes képelemzési eljárások képesek ekkora mennyiségű adatot feldolgozni. Számos matematikai algoritmus létezik, amelyik képes leírni a vizuális információhordozón látható alakzatokat azonban az igazsághoz hozzátartozik, hogy ma még ezek az algoritmusok nem minden esetekben működnek tökéletesen. Szerencsére ezek az algoritmusok és eljárások fejlődnek, így néhány év múlva talán lehetővé válik, hogy a ma még strukturálatlan adathalmaz teljes egészében feldolgozásra kerül és kereshetővé válik.

Irodalomjegyzék

- Aigrain, P., Zhang, H., Petkovic, D. 1996. Content-based representation and retrieval of visual media: A review of the state-of-the-art. *Multimed. Tools Appl.* 3, 3, 179–202. o.
- Barnard, K., Duygulu, P., Forsyth, D., De Freitas, N., Blei, D. M., Jordan, M. I. 2003. Matching words and pictures. *J. Mach. Learn. Res.* 3, 1107–1135. o.
- Barni, M., Pragotti, A., Piva, A. 2005. Image processing for the analysis and conservation of paintings: Opportunities and challenges. *IEEE Signal Process.* vol. 22, 141–144. o.
- Beretti, S., Bimbo, A. D., Vicario, E. 2001. Efficient matching and indexing of graph models in contentbased retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* Vol. 23. 1089–1105. o.
- L. G. Brown: A survey of image registration techniques. *ACM Computing Surveys*, 24. évf. (1992) 4. sz., 325–376. o.
- Carson, C., Belongie, S., Greenspan, H., Malik, J. 2002. Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Trans. Pattern Anal. Mach.* Vol. 8., 1026–1038. o.
- Chen, Y. Wang, J. Z. 2002. A region-based fuzzy feature matching approach to content-based image retrieval. *IEEE Trans. Pattern Anal. Mach.* Vol. 9. 252–1267.
- T.F. Cootes – D Cooper – C.J. Taylor – J. Graham: A trainable method of parametric shape description. *Image and Vision Computing*, 10. évf. (1992), 289–294. o.
- Ritendra Datta: *Image Retrieval: Ideas, Influences, and Trends of the New Age*. The Pennsylvania State University. 2008. 60 oldal. *ACM Computing Surveys*, Vol. 40, No. 2
- Mathias Eitz, Kristian Hildebrand, Tamy Boubekeur és Marc Alexa: *PhotoSketch: A Sketch Based Image Query and Compositing System*. 2009. 4 oldal.
- Jeong, S., Won, C. S., Aangrayr, R. 2004. Image retrieval using color histograms generated by Gauss mixture vector quantization. *Comput. Vision Image Understand.* Vol. 9, 44–66. o.
- Lew, M., Sebe, N., Djeraba, C., Jain, R. 2006. Content-based multimedia information retrieval: State-of-the-art and challenges. *ACM Trans. Multimed. Comput. Commun. Appl.* vol. 1, 1–19. o.
- Jia Li, Wang, J.Z.: *Real-Time Computerized Annotation of Pictures*, *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on* (Volume:30 , Issue: 6)
- Arnold W.M. Smeulders : *Content-Based Image Retrieval at the End of the Early Years*. 2000. 32 oldal. *IEEE Transactions on pattern analysis and machine intelligence*, Vol. 22, No. 12
- C. Grimm, J. J. LaViola Jr. : *A descriptor for large scale image retrieval based on sketched feature lines*. *Eurographics Symposium on Sketch-Based Interfaces and Modeling*. 2009.
- Dr.V.Kannan: *Efficient use of MPEG-7 Color Layout and Edge Histogram Descriptors in CBIR Systems*. *Global Journal of Computer Science and Technology*.
- Sean D. MacArthur, Carla E. Brodley, és Avinash C. Kak: *Interactive Content-Based Image Retrieval Using Relevance Feedback*. 2002. 21 oldal. *Computer Vision and Image Understanding* Vol. 88, No. 55–75 o.
- Pietro Perona – Jitendra Malik: *Scale-space and edge detection using anisotropic diffusion*. In *Proceedings of IEEE Computer Society Workshop on Computer Vision* (konferenciaanyag). 1987. November, 16–22. o.

- Arnold W. M. Smeulders – Marcel Worring – Simone Santini – Amarnath Gupta – Ramesh Jain:
Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal.
Mach. Intell., vol. 22. (2000. December), 1349–1380. o. ISSN 0162-8828 o.
- Yuehua Tao, Youming Xia, Tianwei Xu, Xiaoxiao Chi: Research Progress of the Scale Invariant
Feature Transform (SIFT) Descriptors. Journal of Convergence Information Technology
Volume 5, Number 1, 2010.
- Chee Sun Won, Dong Kwon Park, és Soo-Jun Park:Efficient Use of MPEG-7 Edge Histogram
Descriptor. Korea. 2001. ETRI Journal